Medical Image Analytics for Non-Invasive Lung Cancer Screening

Synopsis of the PhD Thesis

Submitted by

Mr. Pushkar Sathe

(Roll No 70832000001)

Department of Electronics and Telecommunications

Supervisor(s)
Dr. Alka Mahajan
Dr. Mitusha Verma

submitted in partial fulfilment of the requirements of

DOCTOR OF PHILOSOPHY (Engineering)

MUKESH PATEL SCHOOL OF TECHNOLOGY MANAGEMENT & ENGINEERING
SVKM'sNARSEE MONJEE INSTITUTE OF MANAGEMENT STUDIES (Declared as Deemed-to-be University Under Section 3 of the UGC Act, 1956)
V. L. Mehta Road, Vile Parle (West)
MUMBAI –400056
May 2024

TABLE OF CONTENTS

SECTION NO.	TITLE	PAGE NO.
1	ABSTRACT	
2	INTRODUCTION	3
3	OBJECTIVES	4
4	LITERATURE REVIEW	5
	4.1 Abnormality detection and cancer detection	5
	4.2 Cancer sementation	5
	4.3 Cancer grading and early warning	6
5	CONTRIBUTION	7
	5.1 Cancer detection and segmentation	7
	5.2 Cancer grading	23
	5.3 Early warning system	27
6	Conclusion	29
7	Future Scope	30
8	Publications	31
9	References	32

1. ABSTRACT

Lung diseases are seen to be on the rise in the last 30 years. While research has majorly focused on the early detection of lung cancer, effective segmentation, volume estimation and grading of the cancerous nodule are also critical for the treatment of the disease. Further, interpretation of CT scans continues to be a time-consuming manual process with a risk of delayed treatment and possibly subjective evaluation. It is also a sequential process where the CT scans are investigated in the order presented. This may lead to some CT scans belonging to the critical stage getting delayed in the serial process. Also, the cancer detection models classify the scans as cancerous or non-cancerous resulting in a detection system which passes other abnormal scans as normal. So, while the cancerous scans are detected with reasonable accuracy, the doctors end up with examining all the supposedly normal scans to check for other lung diseases. Thus, in addition to detecting the cancerous scans, there is a need to flag other abnormalities to help doctors in prioritizing and ensuring timely action for all other lung diseases too.

Volume estimation and grading of the cancerous nodule depends heavily on accurate segmentation of the nodule. For any automatic lung cancer segmentation system, lung CT scans of hundreds of patients are required along with their corresponding annotated segmentation masks. It is difficult to get accurately annotated data as cancer segmentation of CT scans is done by the radiologists, manually and is also prone to intra and inter-observer variability.

The volume of the tumor plays a very crucial role in deciding the stage of lung cancer which in turn helps in deciding the best treatment and its schedule. Currently used computer-based volume estimation techniques are semi-automatic in nature with limited accuracy. Further, owing to the irregular shape of the cancerous tumor, accurate volume estimation becomes a challenge with regular convolution models.

Selection of the most effective treatment and its schedule depends on the stage of the cancer. There are multiple factors that hold pivotal significance in determining the stage of lung cancer. Volume of the cancerous nodule is one of the major factors in grading the cancerous nodule. Also, the doctors feel a need to have a system which will monitor the periodic follow-up scans and detect the recurrence of the disease or changes in the cancerous nodule after treatment.

While literature reports work done on cancer detection, segmentation, and the very limited work on volume estimation and grading, it has been done in silos. All of these techniques still need human intervention for annotation which is a major time-consuming task for radiologists. This research work thus aims to develop an end-to-end automatic system that includes cancer detection, volumetric segmentation, cancer grading, and an early warning system for flagging of anomalies in the periodic screening. The research is undertaken in collaboration with Nanavati Hospital, Mumbai, and all the models are validated on a real dataset obtained from the hospital. Currently, the cancer detection models focus only on classifying scans as cancerous or normal, bypassing the other abnormalities. A modified two-stage cancer detection system is developed in which the first stage of screening segregates normal and abnormal scans, placing the abnormal scans at the top of the stack. This will ensure prioritized attention and quick treatment. The second stage is designed to further segregate abnormal scans into cancerous and non-cancerous but abnormal scans. The proposed model gave an accuracy of 90.5% for abnormality detection and 99.7% for the cancer detection model.

The scans classified as cancerous are then applied to the segmentation model. The proposed CLT (Cancer-Lymph-Trachea) segmentation model is designed to segment the cancerous nodules, lymph nodes and trachea simultaneously. Segmenting other classes along with cancer is required for grading purposes later. This simultaneous segmentation reduced the model's redundancy and computation time. Three separate annotation masks were generated and used for the training. This allowed the model to learn about the boundaries of classes in a better way leading to improved segmentation. The mathematical model for the modified convolution was designed to capture the multiscale information at non-integral locations within the feature map. This resulted in improving the accuracy of segmentation in terms of well-defined nodule boundaries and the retention of the actual nodule shape. To enhance the model's ability to selectively focus on important features, the focus module was designed and introduced in every layer of CLT segmentation model. It also incorporated the learnable class-wise weighted multipliers to separate three annotation masks from the combined feature map. The class-wise weighted loss function was designed to balance the segmentation performance of all three classes. The CLT segmentation model achieved an average segmentation accuracy of 92.03%. Achieving higher segmentation accuracy was a priority as the results of the rest of the models depended on it.

Further, the mathematical model was developed to estimate the volume of the cancerous nodule based on the CLT segmentation model results. It was found that interpolation techniques were required to further increase the accuracy of volume estimation. GPR (Gaussian Process Regression) technique of interpolation was noted to be the most suitable for irregularly shaped objects like cancer. The proposed volume estimation model gave an accuracy of 94.18% averaged over 58 patients.

The grading model was further developed that considered the required parameters viz; the highest dimension of cancer nodule, reach of cancer to another lung, and reach of cancer to lymph nodes & trachea. The algorithm was designed to check the presence of cancer in another lung by generating the lung mask, locating the mid-portion of the lungs and checking for the spread using an overlapping technique. The highest dimension of the nodule is a dominant factor that decides the spread of the cancer. A mathematical model was thus developed to calculate the highest dimension of the cancerous nodule from the segmentation results giving an accuracy of 93.23%. An algorithm was also developed to check for the overlap of cancer with lymph and trachea. The proposed grading model considered all the important parameters using a direct technique resulting in a state-of-the-art lung cancer grade estimation system.

The grading model was further modified to address the overlap in the grading stages giving clarity to the doctors about the particulars of the anomaly, by raising four relevant red flags. The red flags were designed to give clarity on a particular irregularity in the detected stage, helping the doctors initiate a quick informed treatment plan. Further, in all types of cancer, there is always a risk of recurrence and hence follow-up periodic scanning for the same patient is important. Monitoring the changes in every follow-up scan taken at regular intervals can help detect any new or growing abnormalities indicating possible progression of the disease. The grading model was extended further and an early warning system was developed not only to maintain the history of a patient but also to aid in the early detection of recurrence and new abnormalities.

2. INTRODUCTION

Lung cancer is the most widespread type of cancer and is the second most common cancer after prostate cancer in males and breast cancer in females. The early detection of lung cancer plays an effective role in diagnosis and leads to an early treatment increasing the likelihood of patient survival rate [1]-[2]. Chemotherapy, which is the main treatment for lung cancer requires knowledge of accurate location of the cancer along with its spread along all the 3 axes. Accurate volume estimation is required for determining the stage of the cancer. Volume estimation can be done using 3D or 2D CT scan images. Though 3D data gives more accurate volume estimation it suffers from lack of annotated data required for training and higher complexity of the segmentation algorithms. So generally 2D data is used to estimate the slice wise area of cancer spread and then volume is obtained using these area values. It is very difficult to get accurately annotated data as manual segmentation is usually performed by radiologists. The human intervention leads to errors arising out of fatigue and subjectivity. To overcome this problem, cancer segmentation models based on deep learning were proposed but they require a large dataset along with corresponding annotations of every CT scan for training to get acceptable segmentation accuracy that leads to precise area calculation. Also, majority of cancer segmentation algorithms fail to clearly define the boundaries of cancerous nodules. While many software's are available for calculating volume automatically, they are semi-automatic in nature as they need demarcations of cancerous portions to be done by radiologists in every or at most on alternate slices. The segmentation accuracy plays a crucial role in deciding the accuracy of the overall cancer grading system. The grade of the cancer helps oncologists determine the aggressiveness of the cancer and tailor treatment plans accordingly. It is an important factor in predicting the likely outcome of the disease. Monitoring changes in the grade of the cancer over time can help assess how well the treatment is working. If the grade increases despite treatment, it may indicate that the cancer is becoming more aggressive or resistant to therapy. Similarly, performing timely screening of lung CT scans for the same patient is critical for monitoring disease progression, detecting recurrence, assessing treatment response, identifying new abnormalities and providing long-term surveillance. No work in the literature is found that covers all required parameters to determine the grade of the lung cancer and to highlight abnormalities in the periodic screening using time series analysis. Leveraging Artificial Intelligence for obtaining the grade of lung cancer offers advantages in terms of accuracy, efficiency, standardization, scalability, integration with clinical workflow, and continual improvement.

3. OBJECTIVES

After reviewing the literature and the gaps in the existing models, the following are decided as the primary objectives of this research:

1. To develop a two-stage cancer detection system that performs the abnormality detection in the first stage and segregates the scans as normal and abnormal.

2. To design a multiclass segmentation model that

    - Performs simultaneous segmentation of three classes namely cancer nodules, lymph nodes, and trachea, and uses focus module and modified loss function to balance segmentation performance of three classes by penalizing poor performing class.
    - Improves the accuracy of segmentation in terms of well-defined boundaries and retention of the exact shape of the cancerous nodules by performing convolution at non-integral locations.

3. To design a volume estimation system and improve its accuracy using interpolation techniques.

4. To develop the grading model that decides the grade of lung cancer based on all the parameters as outlined in TNM classification.

5. To design an early warning system that monitors the changes after follow-up periodic scanning and raises concerns about changes in parameters or recurrence of cancer helping the doctors initiate a quick informed treatment plan.

4. LITERATURE REVIEW

4.1 Abnormality detection and cancer detection

In recent times CT scans are dominantly used for cancer detection and a lot of research on developing AI (Artificial Intelligence) based cancer detection algorithms are reported in literature. Matt Daykin et al. used One Class Support Vector Machine (OCSVM) based model to detect lung abnormality[1]. The method, however, does not use complete images but patches of lung CT scans majorly to increase the size of the data set. Automatic feature extraction or use of deep learning techniques is also not explored fully. Other reported abnormality detection algorithms [2]-[3] majorly dealt with cancers other than lung cancer. Itziar Irigoien et al. applied OCC (One Class Classification) to medical data for the abnormality detection of various diseases like breast cancer, liver disorders, leukemia etc. [2]. The authors compared the performance of four different algorithms namely Gaussian, mixtures of Gaussian, Parzen and typicality approach based on their average AUC (Area Under Curve). The paper reported best results with an average AUC of 77.4% with the typicality approach. L.Tarassenko et. al [3] also applied OCC to investigate normality using large number of available mammograms which do not show any evidence of mass-like structures. The recent advancement in deep learning techniques allow for automatically extracting features from the images, thereby improving the overall performance as compared to conventional CAD (Computer Aided Diagnosis). Use of transfer learning techniques is also reported in literature where fixed feature extraction is done using a pre-trained network. Pasquale Ardimento et al. [4] applied three transfer learning models viz; VGG (Visual Geometry Group), Xception, ResNet for feature extraction and combined the results using ensemble architecture to classify the scans as cancerous and noncancerous. Similarly, the state-of-the-art transferable architectures such as VGG-16, VGG-19, GoogLeNet, Inception-V3, ResNet-18, ResNet-50, ResNet-101, InceptionResNet-V2 and 3D multipath VGG like network have also been used for lung cancer feature extraction[5]-[9] and the performance of SVM (Support Vector Machine) and AdaBoostM2 classifier is analyzed on the deep features extracted from publicly available datasets. Further, all the cancer detection models classify the scans as cancerous and non-cancerous. This also indirectly implies that the noncancerous scans have no abnormality and are flagged as normal. This may lead to a false diagnosis with serious implications. So, the scans classified as non-cancerous will have to be manually processed by the radiologists to further segregate the normal scans. Thus, while the aim of any CAD is to detect cancer, it is also equally important to flag other abnormalities to reduce the diagnostic burden.

4.2 Cancer segmentation

Lung nodule segmentation methods can be categorized into traditional approaches and deep learning-based techniques. Traditional methods encompass threshold and region growing methods, clustering methods, active contour models, and mathematical models [10]-[18]. On the other hand, deep learning-based methods for segmentation can be further divided into 2D and 3D segmentation networks [19]. While traditional segmentation methods do not necessitate a substantial amount of labeled data for model training, they heavily rely on human intervention and are more focused in nature. These methods primarily rely on shallow image features such as grayscale and texture. In contrast, 2D segmentation networks utilize 2D convolution to extract features from images. The original Fully Convolutional Neural Network (FCN) [20], has found extensive use in various image segmentation fields but it lacks in global context information which is very important in the case of medical image segmentation. Ronneberger et al. [21] introduced a U-Net network based on the FCN architecture for medical image segmentation. U-Net incorporates both low-resolution and high-resolution information through skip connections, which is particularly advantageous for segmenting medical images with blurred boundaries. In this approach, low-resolution information is used for target identification, while high-resolution information aids in localization of the segmentation. Consequently, U-Net has served as a foundation for several improved algorithms. To further improve segmentation results, some researchers have explored the integration of Atrous Spatial Pyramid Pooling (ASPP) [22] instead of the intermediate or output layer of U-Net. This approach enables the extraction of multi-scale image information through different perceptual fields [23]-[25] but lacks in retention of exact shape. Some 3D

segmentation networks utilize 3D convolution to extract features from volumetric images, allowing them to better capture the spatial relationship between nodules and surrounding tissues compared to 2D segmentation networks. Duo et al. [26] introduced a 3D fully convolutional neural network capable of automatically segmenting the liver and cardiac great vessels. Milletari et al. [27] incorporated residual connections inspired by ResNet [28] and employed 3D convolution in their V-Net architecture for prostate volume segmentation, departing from the 2D convolution used in U-Net. Deepseed [29] proposed 3D-squeeze-and-excitation (SE) networks, incorporating dynamic scaling with cross-entropy loss [30] to address the sample imbalance problem. Their method was evaluated on the LIDC [31] and LUNA16 [32] datasets, demonstrating promising results for lung nodule segmentation. However, it is worth noting that 3D networks suffer from longer training times and may not necessarily outperform 2D networks in terms of evaluation metrics. Zhou et al. [33] proposed UNet++, an extension of U-Net that incorporates dense skip connections to enable multi-scale fusion and feature acquisition at different levels. In [34] the attention module is used with the Unet algorithm having residual connections for fast backpropagation. But model failed to maintain the exact shape of cancerous nodules.

4.3 Cancer grading and early warning

No constructive work either on the grading of lung cancer or flagging of anomalies in timely screening is found in the literature. In [35] grade classification of pulmonary nodules is performed using multiscale DenseNet but these grades dealt with the probability of malignancy indicating how severe the disease is and not the actual TNM classification [36] of lung cancer. Some basic image processing techniques and feature extraction methods are used to find histogram, area, perimeter, and centroid of cancerous nodules and grade is predicted based on this information [37][38]. But all other important parameters required for grading are not covered leading to incomplete analysis. CNN (Convolution Neural Network) is used for feature extraction [39] of cancerous nodule and they are categorized in following into four grades, namely benign, primary lung cancer, malignant, and suspicious malignant.

5. CONTRIBUTION

5.1 Cancer detection and segmentation

Lung cancer is the most widespread type of cancer and is the second most common cancer after prostate cancer in males and breast cancer in females. The early detection of lung cancer plays an effective role in diagnosis and leads to an early treatment increasing the likelihood of patient survival rate [1]-[2]. First image augmentation is performed and features are extracted as explained in the sections to follow.

5.1.1 Image augmentation

A significant obstacle in training deep learning models for various tasks is the abundance of data required. This challenge is particularly pronounced in the field of medicine, where limited access to costly imaging resources or a scarcity of study subjects can hinder progress. Additionally, the requirement of annotations for every cancerous scan puts a limit on data collection. Consequently, there is a growing trend of employing data augmentation techniques in research involving small data sets [40]-[41]. The demand for extensive data in deep learning networks has spurred the development of various strategies. Only those strategies are used in this work that provide a different visual impact compared to an original image. Vertical flipping, horizontal flipping, PCA (Principal Component Analysis) and image overlay techniques are used in our work for data augmentation.

Figure 1 shows the images generated using corresponding data augmentation methods for a sample CT scan. Figure 2(a) shows the original CT scan. The irregularly shaped cancerous nodule is highlighted using the red bounding box. Figures 2(b) and 2(c) show the results of the flipping operation performed on the original CT scan. After vertical and horizontal flipping the position of the cancerous nodule changed, making it appear as a new image for a deep-learning model. Figure 2(d) shows the result obtained by performing Principal Component Analysis (PCA) with 20 components. Similarly, figure 2(e) shows the new image generated using the image overlay technique.

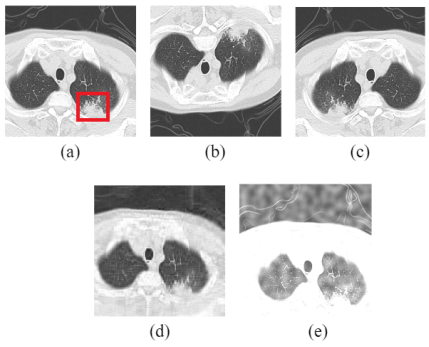


Fig. 1. (a) Original CT scan (b) Vertical flipping (c) Horizontal flipping (d) PCA with 20 principal components (e) image overlay

Image augmentation was performed only on the cancerous scans. A total of 596 cancerous scans (of 97 patients in total) were augmented to 2384. Thereafter, features are extracted using VGG and applied to the abnormality detection and cancer detection models as explained below. In the proposed system, only the first 13 convolutional layers out of the 16 layers of VGG were used for feature extraction as the remaining 3 layers do not contribute to feature extraction. They are used to reduce extracted feature size and hence not applicable here. Figure 2 shows the layered architecture of VGG-16.

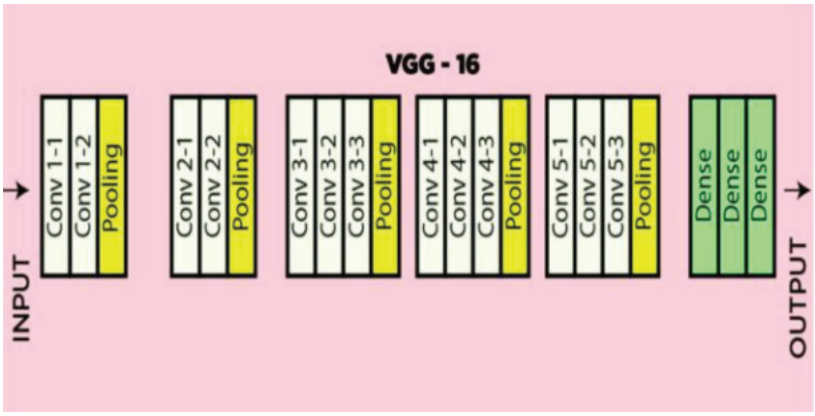
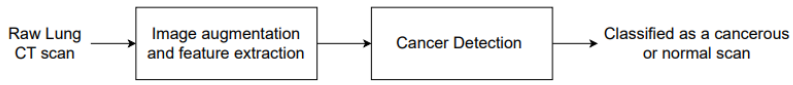


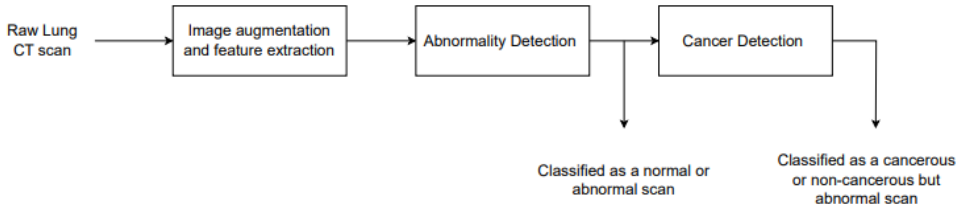
Fig. 2. Layered architecture of VGG-16

5.1.2 Abnormality detection

The Figure 3(a) shows the block diagram of an existing system of cancer detection. Raw CT scans are applied to the image augmentation and feature extraction model and then extracted features are applied to the cancer detection system. Cancer detection model is trained to classify the applied scans as cancerous or normal. So the radiologists need to exclusively assess the segregated normal scans to check for any other abnormality. So we modified the existing system by adding an abnormality detection system block to forming a proposed cancer detection system as shown in the Figure 3(b). In our case, applied scans are applied to the abnormality detection model which is trained to classify applied scans as normal or abnormal. This successfully removes all normal scans from the further checking. Only abnormal scans are applied to the cancer detection model that classifies them as cancerous or non cancerous but abnormal scans giving radiologist a small set to investigate further for any other abnormality.



(a)


Raw Lung CT scan
Image augmentation and feature extraction
Abnormality Detection
Cancer Detection
Classified as a normal or abnormal scan
Classified as a cancerous or non-cancerous but abnormal scan

(b)

Fig. 3.Block diagram of (a) the existing cancer detection system (b) the proposed cancer detection system

In the proposed two stage system, the very first step of prescreening removes all normal scans. Only the identified abnormal scans are applied to the cancer detection model. It segregates the applied scans as cancerous and non-cancerous scans. The rejected lot of this model are flagged off as noncancerous, but abnormal scans which may belong to any other lung abnormality like Tuberculosis, Covid-19, Pneumonia, etc.

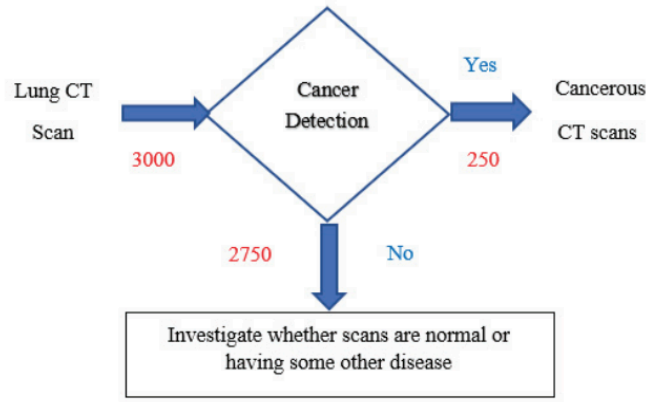


Fig. 4. (a) flowchart of the existing system

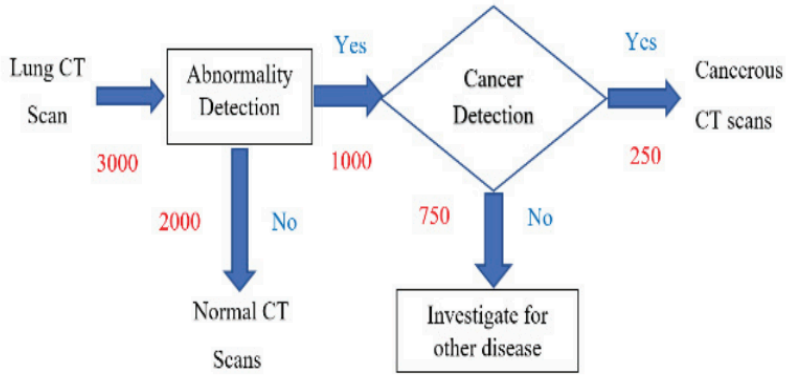


Fig. 4.(b) flowchart of the proposed system

The example given in Figure 4 above indicates that with the proposed system radiologists will need to check just one-fourth of the scans as compared to the current practical scenario where a radiologist manually and sequentially checks scans of approximately a minimum of 50 patients on an average, every day. Further, with the validation dataset, end to end prediction time required for both the systems was found to be almost similar. In one case, all scans are applied to the cancer detection model and in the other case, they get filtered by the first model and only abnormal scans are applied to the cancer detection model. It can thus be concluded that in almost the same time, the proposed system categorizes the input scans as normal, cancerous, and non-cancerous but abnormal cases. While the proposed system uses one additional module for abnormality detection, it needs only normal CT scans for training that are easier to obtain as compared to cancerous scans. Also, there is no need of using data augmentation methods as sufficient normal CT scans are available.

The abnormality detection model was trained using only normal lung CT scans. 3250 normal lung CT scans were used to train the model. The model learned the distribution of normality when many instances of normal

samples were used to train it. Once the model was trained, test samples were applied to see how much they differed from normality to classify them as abnormal and normal. The abnormality detection model not only separated abnormal scans but also helped establish the priority of processing for radiologists. A comparative study was conducted using two classifiers; Isolation Forest and OCSVM (One Class Support Vector Machine). The abnormality detection model was validated on a test dataset of 160 normal and 160 abnormal scans. The classification results of Isolation Forest and OCSVM are summarized in the table 1 part (a) and (b). Table 1 part (c) shows the comparison of evaluation metric values obtained for isolation forest and OCSVM classifiers.

		Predicted class	
		Abnormal	Normal
Actual Class	Abnormal	152	8
	Normal	38	122

(a)

		Predicted class	
		Abnormal	Normal
Actual Class	Abnormal	160	0
	Normal	30	130

(b)

Algorithm / Metric	Accuracy	Precision	Recall	Specificity	F1 - score
Isolation forest	85.62 %	80 %	95 %	76.25 %	86.85 %
OCSVM	**90.62 %**	**84.21 %**	**100 %**	**81.25 %**	**91.42 %**

(c)

Table.1. Classification results of abnormality detection classifiers (a) Confusion matrix for isolation forest classifier (b) Confusion matrix for OCSVM classifier (c) Comparison of evaluation parameters

As the recall (or sensitivity) obtained is 100%, it is clear that all abnormal scans were correctly classified by OCSVM. Also, the AUC plot for each classifier is shown in Figure 5. It is seen that the ROC curve of OCSVM is closer to the ideal ROC curve.

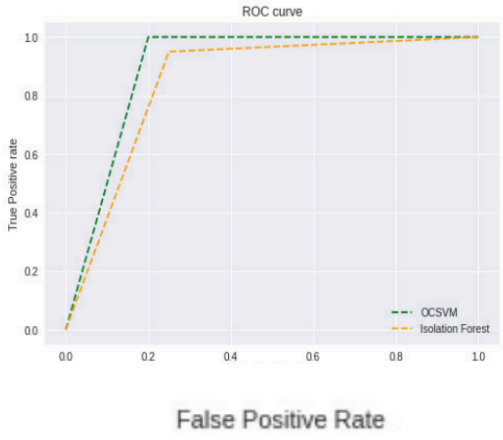


Fig. 5. AUC plot of abnormality detection classifiers

5.1.3 Cancer detection

The CT scans classified as abnormal by the abnormality detection model were applied to the cancer detection model in the second stage. The cancer detection model was trained using both normal and cancerous CT scans. 3250 normal and 2384 cancerous CT scans leading to a total of 5634 scans were used in training. This model was designed to separate cancerous scans from abnormal scans. The extracted features were applied to three different classifiers namely SVM (Support Vector Machine), Decision Tree and Random Forest and their comparative study follows in the sections below. The model was validated on a test dataset of 200 normal and 200 cancerous scans. Tables 2(a), (b) and (c) show confusion matrices of SVM, Decision Tree and Random Forest algorithms respectively. It was found that classification results are optimum for random forest classifier when compared to other classifiers. It's confusion matrix shows that all 200 cancerous scans and 199 scans out of 200 non-cancerous scans were correctly classified. Table 2(d) shows a comparison of evaluation metric values obtained for the three classifiers. Again, the evaluation metrics for random forest classifier outperformed the other two. As all the cancerous scans were correctly classified by Random Forest, recall (or sensitivity) obtained is 100%. The AUC plot (Figure 6) for each classifier shows that the random forest classifier gives the best performance.

		Predicted class	
		Cancerous	Non-cancerous
Actual Class	Cancerous	191	9
	Non-cancerous	6	194

(a)

		Predicted class	
		Cancerous	Non-cancerous
Actual Class	Cancerous	195	5
	Non-cancerous	4	196

(b)

		Predicted class	
		Cancerous	Non-cancerous
Actual Class	Cancerous	200	0
	Non-cancerous	1	199

(c)

Algorithm / Metric	Accuracy	Precision	Recall	Specificity	F1 - score
SVM	96.25 %	96.95 %	95.5 %	97 %	96.22 %
Decision tree	97.75 %	97.98 %	97.5 %	98 %	97.73 %
Random forest	**99.75 %**	**99.5 %**	**100 %**	**99.5 %**	**99.74 %**

(d)

Table 2.Classfication results of cancer detection calssifiers (a) Confusion matrix for SVM classifier (b) Confusion matrix for decision tree classifier (c) Confusion matrix for random forest classifier (d) Comparison of evaluation parameters

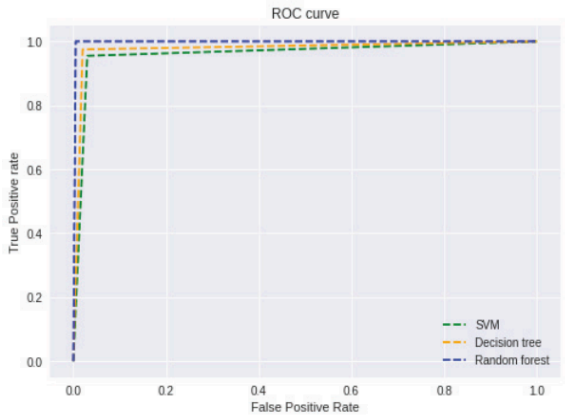


Fig. 6. AUC plot of cancer detection classifiers

Based on the comparative study of classifiers for the two models, OCSVM was used for abnormality detection and Random Forest was used for cancer detection.

5.1.4 Segmentation

The segmentation of medical images is challenging in terms of boundary detection and retention of the exact shape of irregularly shaped objects like cancer. In this work, the multiclass segmentation model is designed to handle the constraints of the irregularly shaped classes namely cancer, lymph and trachea. We applied an image of size (512,512,3) to our proposed segmentation algorithm. Instead of drawing annotations for all 3 classes in the same mask, we generated three different masks for three different classes. We know that, depending on the stage of cancer there is a high chance of overlap of cancer on lymph or trachea. So if we mark annotations for all three classes in the same mask, learning of the model would become difficult because of the overlapping between classes. This may even lead to the poor segmentation of the classes in terms of perfect shape and boundary extraction. We can not take any risk of imperfect segmentation as the grade of the cancer heavily depends upon the highest dimension of cancer and the reach of cancer to the lymph or trachea. So in the training of images of size (512,512,3), we used 3 masks each of size (512,512,1) containing annotation of 1 class each. Also, this simultaneous segmentation led to reduced redundancy and computation time.

The proposed segmentation model is as shown in Figure 7. Various parts of the architecture are explained in the sections to follow.

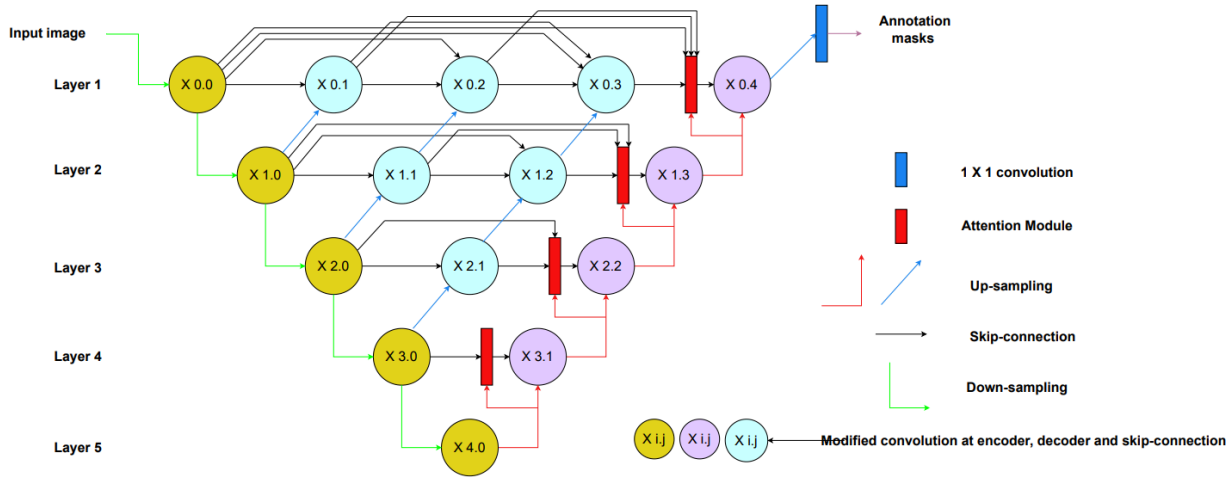


Fig. 7. The proposed segmentation model

a) Encoder

Every encoder layer performs a modified convolution operation on the received feature map, followed by ReLU and max pooling operations. When the input image of size (512,512,3) is applied to the layer1 of encoder, it undergoes convolution operation with padding value 1 and a total of 64 filters. This results in a feature map of the same size as that of input image but with 64 filters. i.e. (512, 512, 64). Then ReLU just replaces all negative values by zero. So the dimension of the tensor remains the same. After this, the max-pooling operation is performed leading to a tensor size of (256, 256, 64) which is then applied to the next encoder layer.

For every next layer of the encoder, the number of filters is doubled than that of the previous layer. Also, the size of the feature map gets halved layer by layer because of the max pooling operation. So outputs of encoder layers 2,3,4 and 5 are (128,128, 128), (64, 64, 256), (32,32, 512) and (16,16, 1024) respectively.

b) Modified Convolution:

The standard convolution grid is as shown in the figure below.

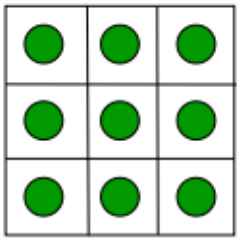


Fig. 8 The standard convolution grid

For any pixel, the weighted addition with a particular convolution mask is performed for surrounding pixels to extract the features

The grid R defines the receptive field size and dilation. For example,

R = {(−1, −1),(−1, 0), . . . ,(0, 1),(1, 1)} defines a 3 × 3 kernel with dilation 1. For each location Po on the output feature map y, we have

$$\mathbf{y}(\mathbf{p}_0) = \sum_{\mathbf{p}_n \in \mathcal{R}} \mathbf{w}(\mathbf{p}_n) \cdot \mathbf{x}(\mathbf{p}_0 + \mathbf{p}_n) \tag{1}$$

where Po Pn enumerates the locations in R.

The main problem with the standard convolution is that the receptive field remains constant regardless of the size of the object being analyzed. However, this approach is not ideal when dealing with objects of varying sizes. In the context of medical image target segmentation, the lesions that need to be segmented often have irregular shapes and sizes. Dilated convolution addresses this by introducing gaps in filter taps, effectively enlarging the field of view without increasing parameters or computation. Though it was successful in capturing multi-scale information to some extent, fixed-location convolution remains a hurdle in semantic segmentation, especially in scenarios like cancer where the object lacks a fixed shape. So we have proposed a modified convolution that allows the receptive field to be dynamically adjusted based on the features of the input data. The process of convolution takes place at some better location than the fixed point as in the case of standard convolution. These offsets are trainable. This modified convolution allows the model to focus more on the relevant regions of the input feature map leading to better segmentation in terms of higher IoU and retention of cancer nodule boundaries. Figure 9(a) shows the standard convolution mask and (b) shows the convolution mask for the modified convolution with dilation rate 2. The arrows indicate the new places at which convolution operation takes place.

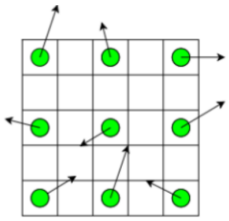


Fig. 9. (a) standard convolution mask (b) Modified convolution mask

In modified convolution, the regular grid R is augmented with offsets $\{\Delta\mathbf{p}_n | n = 1, ..., N\}$, where N = |R| and dilation rate **D**.

$$\mathbf{y}(\mathbf{p}_0) = \sum_{\mathbf{p}_n \in \mathcal{R}} \mathbf{w}(\mathbf{p}_n) \cdot \mathbf{x}(\mathbf{p}_0 + \mathbf{D}\,\mathbf{p}_n + \Delta\mathbf{p}_n) \tag{2}$$

where P denotes the arbitrary location whose pixel value is obtained through the process of bilinear interpolation.

Bilinear interpolation first identifies the surrounding four grid points that are closest to the offset location. These grid points define a local neighborhood. Bilinear interpolation calculates weights for each of the surrounding grid points based on their distances to the offset location. The weights are inversely proportional to the distances, meaning closer grid points have higher weights. Next, the signal values at the surrounding grid points are multiplied by their corresponding weights. These weighted signal values are then summed together to obtain the interpolated value at the offset location. It allows the convolutional kernel to sample input features from non-integer coordinates,

effectively enabling more flexible and adaptive sampling of the input feature map based on the predicted offsets. This adaptive sampling mechanism helps the model to capture fine-grained spatial details and improve its performance in semantic segmentation.

c) Decoder

At decoder layer 4, upsampling of (16,16,1024) takes place to tensor size (32, 32, 1024). From skip connection of layer 4 of encoder, we receive tensor of size (32, 32, 512). These 2 tensors get concatenated and a tensor of size (32,32,1536) is formed. Then after performing modified convolution with 512 filters, it becomes of size (32,32,512) which is handed over to decoder layer 3.

Similarly, the outputs of decoder layers 3,2 and 1 are (64, 64, 256), (128,128, 128) and (256, 256, 64) respectively. The output of decoder layer 1 is then upsampled to (512,512, 64). After 1X1 convolution with 3 filters, the tensor size gets changed to (512, 512,3) indicating three black and white annotated masks one for each class. The sigmoidal activation function is used instead of the softmax activation function. If softmax is used, then based on three probability values, it will try to classify every pixel under a single class only. As in our case, for any given pixel we can parallelly have white pixels in cancer, lymph or trachea annotation masks (we have overlap between classes), we have used the sigmoidal function.

d) Skip pathways

Skip pathways are used to club the features from different levels together and provide spatial information to the decoder to contribute to better segmentation. All the convolutions performed in skip pathways, maintain the dimension of the feature map of that layer intact and only the number of filters gets changed.

X 0,1 receives input from X 0,0 (size 256, 256, 64) and X 1,0 (size 128, 128, 128). X 1,0 is upsampled to size 256, 256, 128. So after concatenation with X 0,0, it becomes of size (256, 256, 192). This after convolution with 64 filters gets changed to the size (256, 256, 64). Finally this tensor is applied to the focus module.

we formulate the skip pathway as follows: let $x^{i,j}$ denote the output of node Xi,j where i indicates the down-sampling layer along the encoder and j indicates the convolution layer along the skip pathway. The stack of feature maps represented by $x^{i,j}$ is computed as

$$x^{i,j} = \begin{cases} \mathcal{H}\left(x^{i-1,j}\right), & j = 0 \\ \mathcal{H}\left(\left[\left[x^{i,k}\right]_{k=0}^{j-1}, \mathcal{U}(x^{i+1,j-1})\right]\right), & j > 0 \end{cases} \tag{3}$$

where function H(·) is a convolution operation followed by an activation function, U(·) denotes an up-sampling layer, and [ ] denotes the concatenation layer. Basically, nodes at level j = 0 receive only one input from the previous layer of the encoder and nodes at level j = 1 receives two inputs, both from the encoder sub-network but at two consecutive levels; and nodes at level j > 1 receive (j + 1) inputs, of which j inputs are the outputs of the previous j nodes in the same skip pathway and the last input is the up-sampled output from the lower skip pathway.

e) focus module

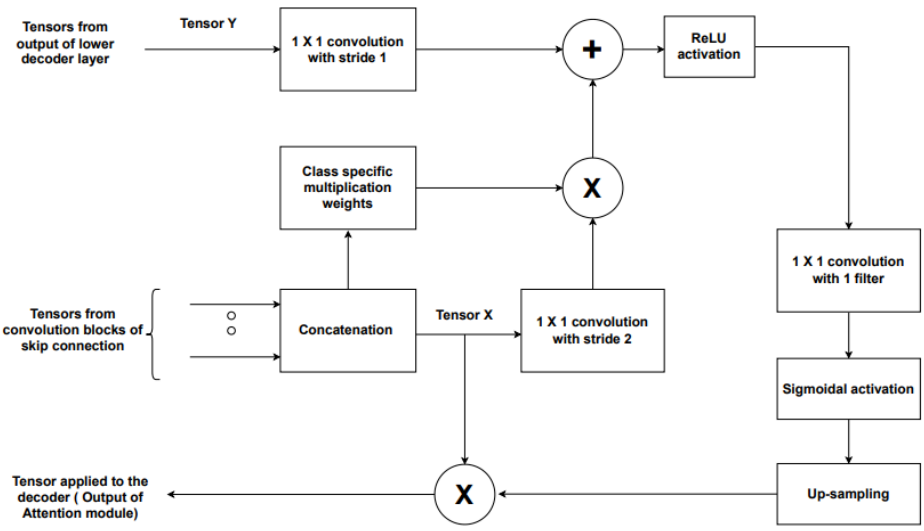


Fig. 90. focus module

The focus module shown in Figure 10 is associated with the skip connection of every layer and its output is applied to the decoder of the corresponding layer. It receives a feature map tensor named 'X' from the single or multiple convolutional blocks of the corresponding skip connection and another tensor named 'Y' from the lower decoder layer. In case of upper layers, it has multiple inputs coming from all previous blocks. They are concatenated first forming tensor X. This tensor X carries spatial information whereas tensor Y carries feature information. These two tensors are converted to the same dimension and then added as shown in the figure. So resulting addition will be higher for a particular pixel only if both the feature maps are strong at that pixel. So on an example basis, in the case of layer 1, the focus module receives tensor X of size (256, 256, 256) which is the concatenation of 4 tensors each of size (256, 256,64) coming from each convolution block of skip connection of layer 1 and tensor Y of size (128, 128, 128). 1X1 convolution is performed on X with stride 2 and 64 filters that produce tensor of size (128, 128, 64). Parallelly tensor Y undergoes 1X1 convolution with stride 1 and 64 filters giving tensor of size (128,128,64) at its output as shown in the figure. Now tensors X and Y are of the same size and so, are added. Then this addition signal undergoes ReLU followed by 1X1 convolution with a single filter leading to a tensor of size (128,128,1). These are the weight values showing importance of spatial and features information obtained using addition. These weight values are supposed to be in the range of 0 to1 instead of 0 to infinite as currently. So sigmoidal activation function is applied on all these weights to bring them in the 0 to 1 range, 1 being more important. Then this tensor is upsampled to the size (256,256,1). Finally the multiplication of this tensor with tensor X is performed which is of size (256,256,64) and new tensor of size (256, 256, 64) is generated which is applied to decoder of layer 1. This tensor holds much better feature map as compared to the tensor X as this feature map has got scaled to smaller or larger values based on the combination of spatial and feature informations.

As we have used separate masks for the three classes, we use 3 focus modules in the uppermost layer. In lower layer focus modules class specific multiplication weights are set to 1 making them inactive. For focus module of the layer 1, these weights are trainable and model adjsuts their values depending on the class to be segmented. The concatenation signal holds the combination of spatial and feature-level information making it capable of deciding class specific multiplication weights. These 3 focus modules apply results to 3 decoders of layer 1 finally leading to 3 tensors each of size (512,512,1) representing segmented masks of each class.

Mathematically, an equation of the focus module can be represented as

$$X = X \,.\, (\boldsymbol{\sigma} [X \,.\, W + Y]) \quad (4)$$

where W indicates the class-specific multiplication weights and $\boldsymbol{\sigma}$ indicates the sigmoidal function.
So it is clear from the equation that tensor X gets multiplied by the focus factor ($\boldsymbol{\sigma}$ [X . W + Y]) which depends on spatial information (tensor X), feature information (tensor Y) and class-specific multiplication weights (W). Pixels at which X and Y both are stronger, they are considered as important pixels from a segmentation point of view and their focus factor is adjusted to the higher value. Class-specific multiplication factor further helps in making the model learn about class-specific segmentation to achieve 3 masks one for each class separately.

The proposed algorithm provides the segmentation approach that uses not only modified convolution to club features from different resolutions but also an focus module that scales the feature map based on spatial and feature information leading to more accurate segmentation. Our proposed segmentation approach has successfully overcome the problem of less training data.

f) Loss function used in training :

The loss function quantifies the discrepancy between the predicted output of the model and the ground true values. The primary goal during the training phase of a deep learning model is to minimize the loss function. By minimizing the loss function, the model learns to make better predictions and improve its performance on the segmentation task.

For a mask of any class, we expect to have the foreground white and the background black. Binary cross-entropy loss can be problematic in scenarios with a severe class imbalance between foreground and background classes, such as in semantic segmentation tasks where background pixels significantly outnumber foreground pixels. In such cases, the model may focus more on correctly classifying the abundant background pixels since they contribute more to the overall loss calculation. As a result, the loss value can become dominated by the background class, potentially leading to misguidance in the training process as we mainly aim to have a perfect classification of the foreground class.

In our work, we used the loss function as a combination of focal loss and dice loss instead of binary cross-entropy loss. The focal loss is beneficial to address the class imbalance issue. Focal loss is just an extension of the cross-entropy loss function that would down-weight easy segmentation portion and assign more weightage to hard segmentation portion i.e. foreground pixels. Focal Loss introduces a "focusing parameter" that down-weights the loss assigned to well-classified examples. So Focal Loss reduces the loss contribution from easy examples and increases the importance of correcting misclassified examples. In addition to the focusing parameter, Focal Loss includes a "modulation factor" that balances the contribution of positive and negative examples to the loss.

The Focal Loss can be represented as:

$$FL = -\alpha t(1- pt)^{\gamma} \log \log(pt) \text{ summed over all values of pt} \quad (5)$$

Where pt is the predicted probability of the true class
αt is the balancing factor which is set to α for foreground and $(1-\alpha)$ for background to down-weight loss obtained from well-classified examples. α is a hyperparameter learning during the training.
γ is the focusing parameter that focuses on the training of hard examples, reducing the influence of easily classified examples.

The dice Loss is a metric used in segmentation tasks where the actual delineation of the region of interest is essential for decision-making. It is well suited for applications involving irregular structures like cancer.

Mathematically it is defined as given in the equation

$$DL = (2 * |X \cap Y|) / (|X| + |Y|) \quad (6)$$

Here, X and Y represent the sets or regions of pixels that belong to the predicted segmentation and the ground truth segmentation, respectively.

So the loss function used for a single class looks like

$$\text{Total Loss (TL)} = FL + DL \quad (7)$$

The Final Loss Function (FLF) should combine the total loss obtained from all the 3 classes separately. When we blindly added all three total losses, the class offering higher loss (cancer class in our case) was balanced by the class with lower loss (the trachea class). As the main aim of the model during training is to minimize FLF, the poor-performing class is not getting separately penalized and the model continues to perform poorly for that particular class. To avoid this, the FLF is obtained as a weighted total loss from all 3 classes as shown below.

$$FLF = \sum_{i=1 \text{ to } 3} (W_i) \times (TL_i) \quad (8)$$

Where W_i and TL_i are the weight value and the total loss associated with the i^{th} class.
The weight values are obtained as

$$W_i = K \times TL_i \quad (9)$$

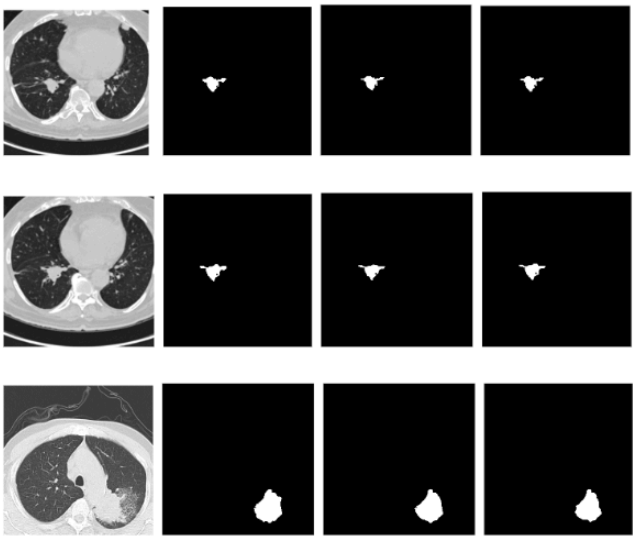
Where K is the controlling parameter that determines how severely we want to reduce the imbalance between the three classes. We have used K=3.
So, the higher the TL (Total Loss) of the particular class, the higher the corresponding weight value and so higher the total loss portion contributed by that class. In this scenario, the model pays higher priority to that class and aims to reduce the total loss of the corresponding class, which is indeed desired.

In this work, the Intersection over Union (IoU), Dice Similarity Coefficient (DSC), Relative Volume Difference (RVD), Average Surface Distance (ASD) and Hausdorff Distance (HD) were used to evaluate the image segmentation results. Results of segmentation of three classes using proposed algorithm are as discussed below. The performance metrics for 3 patients CTP1, CTP2 and CTP3 for cancer segmentation are highlighted in Table 3.

Patients	Algorithm/parameters	IoU	DSC	RVD	ASD	HD
CTP1	Unet++	0.8751	0.933	0.1426	0.1554	2.8284
	CLT Segmentation model	**0.945**	**0.972**	**0.0582**	**0.0656**	**2.2361**
CTP2	Unet++	0.8754	0.934	0.1423	0.2546	6.7082
	CLT Segmentation model	**0.9287**	**0.963**	**0.0767**	**0.123**	**4.4721**
CTP3	Unet++	0.8698	0.93	0.1496	0.1747	2.8284
	CLT Segmentation model	**0.9368**	**0.967**	**0.0674**	**0.0817**	**2.00**
Average for 58 patients	Unet++	0.8716	0.9152	0.1446	0.1982	2.828
	CLT Segmentation model	**0.9209**	**0.9586**	**0.0682**	**0.0972**	**2.2471**

Table 3. comparison of various segmentation metrics for Unet++ and the proposed segmentation algorithm

Figure 11 shows the results of segmentation obtained using U-net++ and the proposed algorithm for patients. It is seen clearly that the proposed algorithm retains the shape of a nodule by correctly demarcating its boundaries and provides more accurate segmentation compared to U-net++. The spider chart showing the comparison of U-net++ and the proposed algorithm for various segmentation metrics is shown in Figure 12. The larger difference in the areas covered by the two algorithms in the spider chart indicates the difference in their performance.



(a) (b) (c) (d)

Fig. 11. (a) Original CT scan (b) Ground truth (c) result of U-net++ (d) result of proposed algorithm

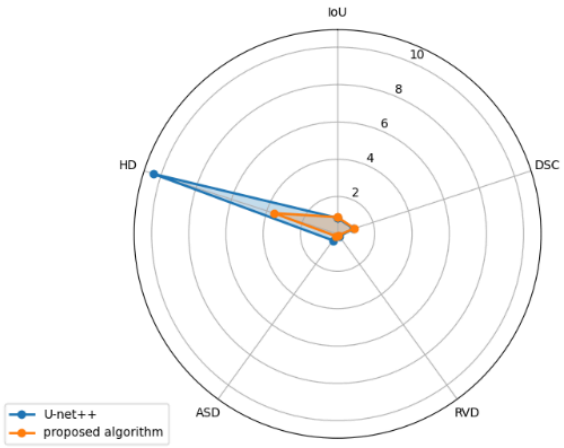


Fig. 12. The spider chart of comparison of U-net++ with proposed algorithm.

The average value of segmentation, an IoU of 0.9035, obtained with the proposed segmentation algorithm is nearly 5% higher than that given by the U-net++ algorithm.

g) Volumetric estimation

When a series of CT scans of the patient is applied to the cancer segmentation model, it provides corresponding masks having cancerous portions highlighted for every CT scan. These annotated masks are applied to the volume estimation model that calculates the area of the cancerous portion of every scan and uses the interpolation method to find the volume of the tumor. From the segmented mask, a number of white pixels are obtained. The area of the cancerous portion is calculated by multiplying the number of white pixels by the area of each pixel which is obtained from pixel height and pixel width functions of the Python library. Likewise, the area is calculated for all consecutive CT scans and then the volume is obtained from these area values using the interpolation method. Figure13 shows the consecutive cancerous scans for one sample patient case. Corresponding area values are 0.435, 2.966, 4.104, 3.8 and 1.755 respectively.

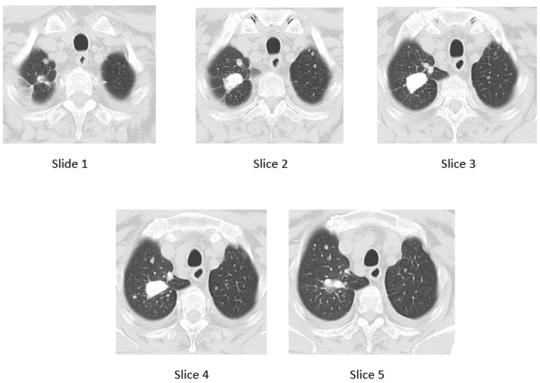


Fig. 13. Consecutive cancerous scans of a sample patient

Each slice is 5 mm thick. This means the real area of cancer spread is known only at intervals of 5 mm each and not in between. Considering the same cancerous area throughout the slice thickness will lead to wrong volume calculation. For better volume estimation, interpolation is used to predict the area values of the cancerous portion 0.01 mm apart slice-wise. Various types of interpolations like linear, bilinear, cubic, Lagrange, GPR were tried to determine the volume of cancer, and their results are compared with the actual volumes obtained from radiologists. GPR interpolation is found to be the best approach.

h) GPR interpolation

The basic idea of GPR is to model the relationship between the input data and the output values as a Gaussian process. A Gaussian process is a collection of random variables, any finite number of which have a joint Gaussian distribution. In GPR, we assume that the output values follow a Gaussian process with a mean function $\mu(x)$ and a covariance function $k(x, x')$ that captures the similarity between given data points and interpolation points. The most commonly used covariance function is the Radial Basis Function (RBF) kernel. This interpolation is the most suitable for irregular data as it can better capture underlying trends and patterns. This property makes GPR a most suitable interpolation method in cancer volume estimation. The key idea of GPR interpolation is to use Bayesian inference to compute the posterior distribution over the unknown value $\hat{y}$ conditioned on the observed data. The posterior distribution is a Gaussian distribution with a mean $\hat{\mu}$ and a covariance $\hat{\Sigma}$. The mean $\hat{\mu}$ represents the estimated value of $\hat{y}$, and the covariance $\hat{\Sigma}$ quantifies the uncertainty associated with the estimation.

Mathematically GPR interpolation can be derived as follows:
First, the mean function $\mu(x)$ and the covariance function $k(x, x')$ were defined. Then the covariance matrix $K(X, X)$ between the observed input points in X and the covariance vector $k(X, \hat{x})$ between the observed input points in X and the new input point $\hat{x}$ were computed. The covariance scalar $k(\hat{x}, \hat{x})$ between the new input point $\hat{x}$ and itself was then calculated. Finally, the mean vector $\hat{\mu}$ and the covariance matrix $\hat{\Sigma}$ of the posterior distribution were obtained using the formulae [31] given in equations (10) and (11)

$$\hat{\mu} = k(X, \hat{x})^T [K(X, X) + \sigma^2 I]^{-1} y \quad (10)$$

$$\hat{\Sigma} = k(\hat{x}, \hat{x}) - k(X, \hat{x})^T [K(X, X) + \sigma^2 I]^{-1} k(X, \hat{x}) \quad (11)$$

where σ^2 is the noise variance parameter and I is the identity matrix.

The estimated output value $\hat{y}$ at the new input point $\hat{x}$ was obtained by the mean $\hat{\mu}$, and the uncertainty associated with the estimation was quantified by the covariance $\hat{\Sigma}$.

Various types of interpolations like linear, bilinear, cubic, Lagrange, GPR (Gaussian Process Regression) are checked in this work and their results are compared. Table 4 shows the results of these interpolation methods applied to CT scans of 3 patients. VP1, VP2 and VP3 are the volumes estimated for 3 patients P1, P2 and P3 respectively. These volumes are compared with ground truth volumes calculated by the radiologists. It is clearly seen that there is a huge difference in calculated volume and ground truth volume when interpolation is not used. In the case of VP3, volume estimation accuracies with quadratic, cubic, spine, lagrange and GPR interpolations are 88.13%, 88.91%, 90.27%, 91.63% and 94.36% respectively. Similarly, for the other patients it is found that GPR gives the best volume estimation accuracy with an average value of 94.18%.

Nodule volume estimation / Type of interpolation	Ground truth volume	Volume without interpolation	Quadratic	Cubic	Spine	Lagrange	GPR	Volume estimation accuracy in % for GPR
VP1	5.79	6.53	6.37	6.36	6.39	6.35	6.192	93.06
VP2	4.49	4.93	4.13	4.09	4.12	4.18	4.21	93.77
VP3	5.14	5.97	5.75	5.71	5.64	5.57	5.43	94.36
Average for 32 patients	5.17	5.82	5.78	5.81	5.69	5.63	5.47	94.18

Table 4. volume estimations (in cubic cm) for various interpolation methods

Nodule volume estimation	Ground truth volume	Volume :GPR interpolation on U-net++	Volume accuracy in % for GPR on Unet++	Volume :GPR interpolation on proposed algorithm	Volume accuracy in % for GPR on proposed algorithm
VP1	5.79	6.422	89.08	6.192	93.06
VP2	4.49	4.039	89.95	4.21	93.77
VP3	5.14	5.62	90.66	5.43	94.36
Average for 32 patients	5.17	5.66	90.51	5.47	94.18

Table 5. Comparison of volume estimations (in cubic cm) for U-net++ and proposed algorithm

For comparison purpose, we applied GPR interpolation on the segmentation masks obtained using the U-net++ algorithm. Table 5 indicates volumes obtained using GPR interpolation applied to the Unet++ algorithm and our proposed algorithm for VP1, VP2 and VP3. It is seen that the accuracy of the proposed algorithm is approximately 4% higher than that of U-net++. For all volumes VP1, VP2 and VP3 as well as for average it is seen that the volume estimated using the proposed algorithm is closer to the ground truth compared to that of the U-net++ algorithm.

5.2 Cancer Grading

Accurate grading is required for the correct treatment of lung cancer in terms of dosage and duration. The figure 14 below shows the factors on which grade of the lung cancer majorly depends upon.

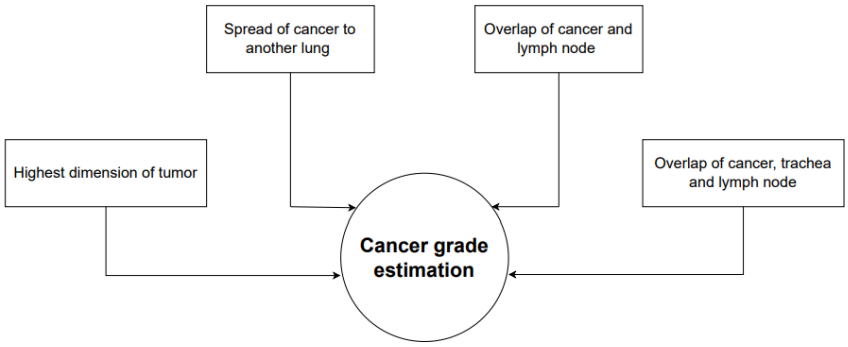

Spread of cancer to another lung
Overlap of cancer and lymph node
Highest dimension of tumor
Overlap of cancer, trachea and lymph node
Cancer grade estimation

Fig. 14. Lung cancer grade estimation factors

5.2.1. Highest dimension estimation

For every slice, the width and breadth of the cancerous portion are calculated in Python from its corresponding annotated mask. The height of the tumor is calculated for the stack of slices put one above the other. The largest dimension of the tumor is one of the important parameters taken into consideration for deciding the grade of lung cancer.

The table 6 shows the comaprison of highest dimension values obtained using standard unet++ and that by the proposed model.

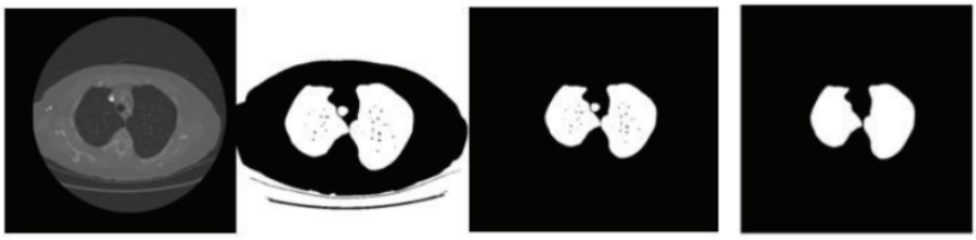
Nodule highest dimension estimation	Ground truth value of highest dimension (cm)	U-net++ with standard convolution		Proposed algorithm	
		Value (cm)	% accuracy	Value (cm)	% accuracy
P1	3.27	2.96	90.52	3.13	95.72
P2	2.49	2.08	83.54	2.271	91.17
P3	2.78	2.492	89.65	2.54	91.37
Average for 74 patients	2.821	2.506	88.84	2.63	93.23

Table 6. Comparison of highest dimension estimations (in cm) for U-net++ with standard convolution and proposed algorithm

Though cancer grading depends on multiple factors, highest dimension of the tumor is considered to be the most important factor. So if that goes wrong, it may lead to the inaccurate prediction of the stage of the cancer and hence the corresponding dosage and the treatment. A similar case is observed in the case of patient P1 of Table 6. The ground truth value of the highest dimension is 3.27 cm. It belongs to the cancer stage 2A as per the TNM classification[32] as it is in the range of '3 to 5 cm'. With the original U-net++, the obtained dimension value is 2.96. Though it is 90% accurate it falls in the category of 'less than 3 cm' and hence belongs to cancer stage 1. This scenario will lead to the wrong prediction and incorrect treatment. On the other hand, the highest dimension value obtained by the proposed algorithm is 3.13 cm and it correctly belongs to stage 2A.

5.2.2. Check of spread of cancer to another lung

Though grading mainly focuses on the characteristics of the primary tumor, one factor deciding grade still requires the presence of the secondary nodules to be checked in the other lung. The annotated mask obtained from the cancer segmentation module is checked against the masked lung portion for overlap. The detailed process is summarised below. Figure 15 shows the results of the same. Initially, the raw DICOM lung CT scan as shown in Figure 15(a) was converted into Hounsfield units (HU), a standardized quantitative scale for representing radio density. Each tissue type possesses a distinct HU range, which remains consistent across different individuals. Lung tissues typically fall within the range of -400 to -600 HU. Subsequently, the image was binarized using a threshold value of -600, as depicted in Figure 15(b). Boundary clearing was then conducted on Figure 15(b) to isolate the lung area, resulting in Figure 15(c). Boundary clearing includes edge detection and noise reduction using median filtering. To further refine the image, an erosion followed by dilation process was implemented to eliminate pixels on the object boundaries morphologically. The complete lung portion demarcated in white is shown in Figure 15(d). Then overlap is checked between this lung portion and the annotation mask obtained from the cancer segmentation model. The ANDing operation is performed to obtain an overlap image. If it has white demarcation in both lungs, it indicates that cancer has reached another lung as well.



(a) (b) (c) (d)

Fig. 15. (a) raw DICOM CT scan (b) binarized image (c) boundary cleared image (d) result of erosion and dilation

5.2.3. To check overlap between cancer and lymph node

In the lung, several groups of lymph nodes play crucial roles in the immune system and lymphatic drainage. Lymph nodes contain immune cells such as lymphocytes (B cells and T cells) and macrophages, which play crucial roles in

recognizing and attacking pathogens (such as bacteria, viruses, and other foreign substances) that enter the lungs. Lymph nodes also act as filters for lymphatic fluid, removing pathogens and cellular debris before returning the fluid to the bloodstream. They help prevent the spread of infection and facilitate the removal of waste products from the lung tissue.

When cancer cells spread to nearby lymph nodes, it signifies a more advanced stage of cancer and can have several harmful implications requiring additional surgery to remove affected lymph nodes or radiation therapy to target cancer cells in the lymph nodes. The risk of metastasis increases when the cancer cells spread up to the lymph nodes as cancer cells can travel through the lymphatic system to reach distant parts of the body. Involvement of lymph nodes by cancer can lead to complications such as lymphedema, impaired lymphatic drainage, and compromised immune function, further impacting the patient's quality of life and overall health. So it is of high importance to find whether cancer cells have reached lymph nodes and especially tracheobronchial lymph nodes to decide on grade and associated further treatment of the cancer.

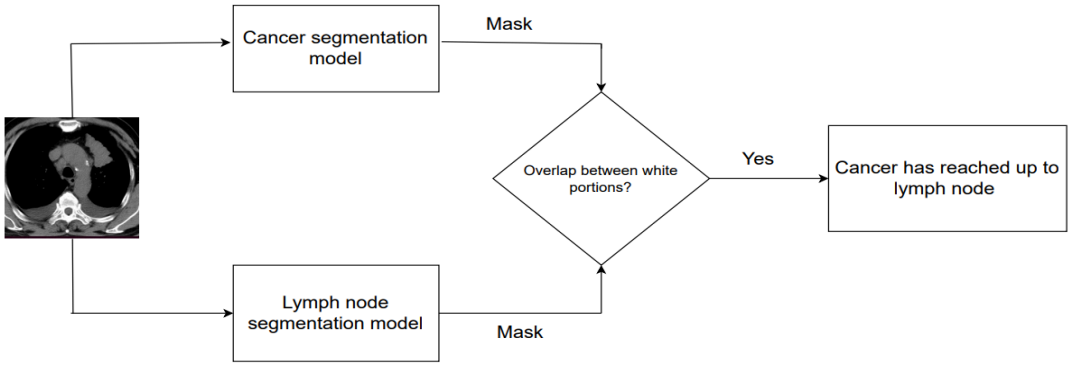


Fig.16. Block diagram of lymph node detection

Our segmentation algorithm provides separate annotated masks of cancer, trachea and lymph. So by ANDing method, we found overlap between classes as shown in Figure 16. The results of lymph node segmentation for sample cases are shown in Figure 17 and corresponding IoU values are specified in table 7.

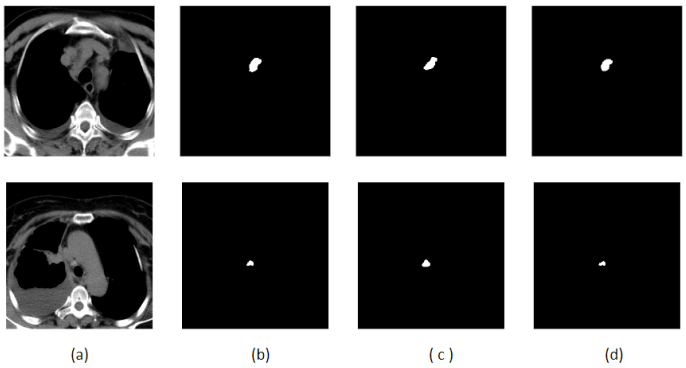


Fig. 17.(a) Original image (b) ground truth (c) segmentation results of Unet++ (d) Segmentation results of the proposed model

Parameter	IoU with standard Unet++	IoU with CLT segmentation model
Patient 1	0.683	0.822
Patient 2	0.762	0.883
Average for 188 samples	0.698	0.864

Table 7. Comparison of IoU values for standard Unet++ versus for the proposed model for lymph node segmentation (average for 188 samples)

5.2.4. To check overlap between cancer, lymph node and trachea

The trachea, also known as the windpipe, is the main airway that carries air to and from the lungs. When lung cancer reaches the trachea, it can physically obstruct the airway, leading to difficulty breathing, shortness of breath, wheezing, and potentially life-threatening respiratory distress. This obstruction can significantly impair the exchange of oxygen and carbon dioxide, leading to hypoxia (low oxygen levels) and hypercapnia (high carbon dioxide levels) in the blood. Lung cancer reaching the trachea often indicates advanced disease progression and metastasis (spread) beyond the lungs.

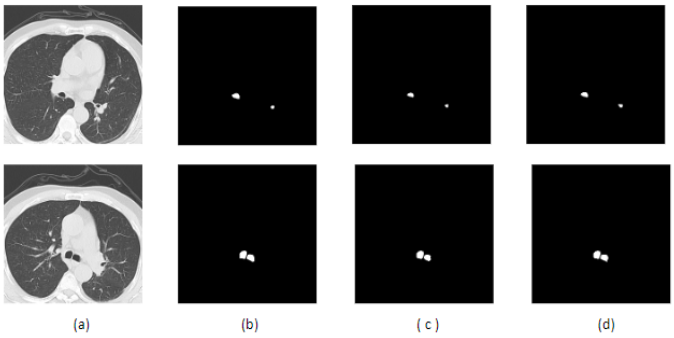


Fig. 18.(a) Original image (b) ground truth (c) segmentation results of Unet++ (d) Segmentation results of the proposed model

Parameter	IoU with standard Unet++	IoU with CLT segmentation model
Patient 1	0.722	0.854
Patient 2	0.903	0.95
Average for 107 samples	0.811	0.916

Table 8. Comparison of Unet++ results for different types of convolutions for trachea segmentation (average of 54 samples)

The results of the trachea segmentation for sample cases are shown in Figure 18 and corresponding IoU values are specified in table 8. It is clearly seen that performance of the proposed model is much better than standard unet++.

5.3 Early warning system

In case of lung cancer, there is alwys a risk of recurrence and so performing timely screening of lung CT scans for the same patient is important. For individuals at high risk of lung cancer, such as current or former smokers, periodic lung CT scans allow healthcare providers to monitor changes in the lungs over time. Comparing scans taken at different intervals can help detect any new or growing abnormalities, indicating possible progression of disease. It helps in early detection of recurrence and an identification of new abnormalities. The CT scans belonging to periodic screening for a sample patient are shown in Figure 19.

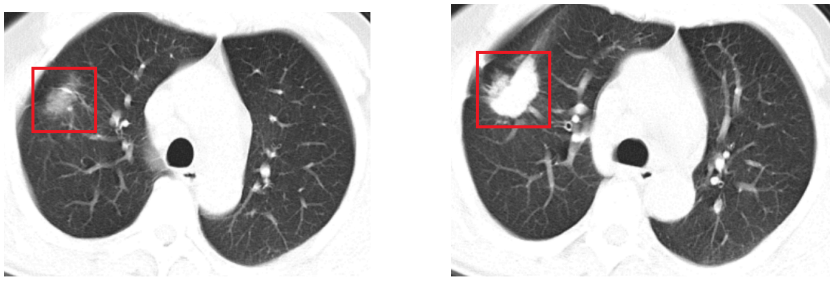


Fig. 19. CT scans belonging to timely screening for a sample patient

We have modified the grading model that incorporates the red flagging in periodic screening. Four red flags are defines as follows
i) RF1 - distance between trachea and cancer is less than 5 cm
ii) RF2 - Cancer spread to lymph and $3 < H < 5$
iii) RF3 - Cancer spread to lymph and $5 < H < 7$
iv) RF4 - $H > 7$

At present, there is no automatic system for monitoring changes and raising a concern after follow-up periodic scanning. While the follow-up may not indicate a change in stage of the cancer but a change in the volume or spread may go undetected. The proposed grading model was further modified to develop an early warning system that monitors the periodic CT scans for changes in parameters and recurrence of cancer. Four relevant warning flags were designed to give clarity on a particular irregularity in the detected stage. The Figure 20 shows the Flowchart of early warning system.

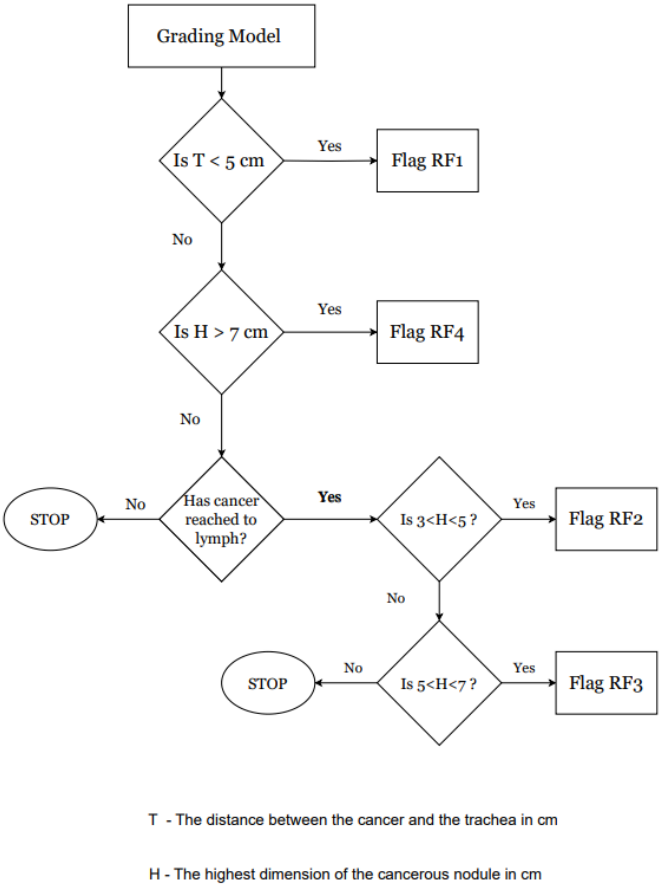


Fig. 20 Flowchart of early warning system.

In summary, we have developed an end-to end system that performs abnormality detection, cancer detection, cancer segmentation, voume estimation, Cancer grading and early warning as shown in the schematic below. (Figure 21)

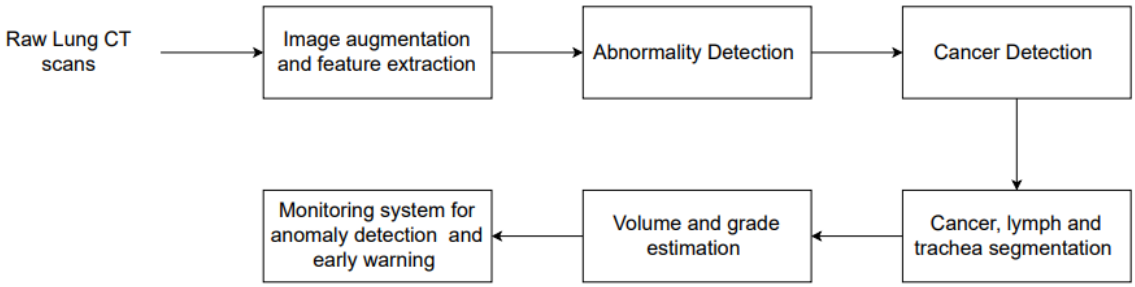

Raw Lung CT scans
Image augmentation and feature extraction
Abnormality Detection
Cancer Detection
Cancer, lymph and trachea segmentation
Volume and grade estimation
Monitoring system for anomaly detection and early warning

Fig. 21. the schematic of the proposed system

Conclusion

Over the past three decades, the prevalence of lung diseases has been increasing. While significant research has been dedicated to the early detection of lung cancer, accurate segmentation, volume estimation, and grading of cancerous nodules are equally critical for effective disease management. Presently, interpreting CT scans remains a laborious manual task, susceptible to delayed treatment and subjective evaluation. The sequential nature of this process may cause critical scans to be overlooked, potentially leading to delays in diagnosis and treatment. Furthermore, existing cancer detection models primarily focus on classifying scans as either cancerous or non-cancerous, overlooking other abnormalities. Consequently, while cancerous scans are identified with reasonable accuracy, doctors must still manually examine supposedly normal scans for other lung diseases. Therefore, there is a pressing need not only to detect cancerous scans but also to flag other abnormalities, facilitating prioritization and timely intervention for all lung diseases.

Accurate volume estimation and grading of cancerous nodules rely heavily on precise segmentation. However, obtaining accurately annotated data for lung cancer segmentation is challenging due to manual annotation by radiologists and inherent variability. Additionally, current computer-based volume estimation techniques are semi-automatic and lack accuracy, particularly for irregularly shaped tumors.

Treatment selection and scheduling depend on the cancer stage, which is influenced by various factors, including nodule volume. Moreover, there is a need for a system to monitor periodic follow-up scans for disease recurrence or changes in nodules post-treatment.

Existing literature on cancer detection, segmentation, and volume estimation is fragmented, often requiring human intervention for annotation, which is time-consuming. This research aims to develop an end-to-end automatic system encompassing cancer detection, volumetric segmentation, grading, and an early warning system for anomaly detection during periodic screening. The research, conducted in collaboration with Nanavati Hospital, Mumbai, validates all models using real-world datasets.

A modified two-stage cancer detection system is proposed to prioritize abnormal scans for prompt attention. The proposed model achieves high accuracy for abnormality and cancer detection. Subsequently, a Cancer-Lymph-Trachea (CLT) segmentation model is developed to simultaneously segment cancerous nodules, lymph nodes, and the trachea, improving segmentation accuracy. Volume estimation and grading models are also developed, providing accurate assessments of nodule characteristics and cancer spread. Furthermore, the grading model is enhanced to identify specific irregularities and facilitate informed treatment decisions. Finally, an early warning system is devised to monitor patients' history and detect recurrence or new abnormalities, ensuring proactive intervention.

Future scope

1) The lung cancer has variety of types. The classification algorithm can be developedfor different types of lung cancers.
2) Growth rate of cancer is very important factor deciding overall spread. The model can be designed to predict the growth rate of the cancerous nodule.
3) Extending the developed system to other type of cancers

Publications

1) Pushkar Sathe, Alka Mahajan, Dr. Deepak Patkar, Dr. Mitusha Verma, “A Two-Stage Computer Aided Diagnostic System for Lung Cancer Detection with Pre-Screening of Other Abnormalities”, International Conference On Futuristic Technologies, Nov. 2022.

2) Pushkar Sathe, Alka Mahajan, Dr. Deepak Patkar, Dr. Mitusha Verma, “Lung cancer grade estimation using deep learning”, IEEE Access. (Under Review)

3) Pushkar Sathe, Alka Mahajan, Dr. Deepak Patkar, Dr. Mitusha Verma, “A Modified Architecture for Lung Cancer Segmentation using Deep Learning” , Journal of Cancer Research. (Under Review)

4) Pushkar Sathe, Alka Mahajan, Dr. Deepak Patkar, Dr. Mitusha Verma, “End-to-End Fully Automated Lung Cancer Volume Estimation System”, International Journal of Electrical and Computer Engineering Systems. (Under Review)

5) Pushkar Sathe, Alka Mahajan, Dr. Deepak Patkar, Dr. Mitusha Verma, “Fully Automated Lung Cancer Highest Dimension Estimation System”, Indian Journal of Cancer. (Under Review)

References

[1] M. Daykin, M. Sellathurai, and I. Poole, “A comparison of unsupervised abnormality detection methods for interstitial lung disease,” Commun. Comput. Inf. Sci., vol. 894, pp. 287–298, 2018, doi: 10.1007/978-3-319-95921-4_27.

[2] I. Irigoien, B. Sierra, and C. Arenas, “Towards application of one-class classification methods to medical data,” Sci. World J., vol. 2014, 2014, doi: 10.1155/2014/730712.

[3] L. Tarassenko, P. Hayton, N. Cerneaz, and M. Brady, “Novelty detection for the identification of masses in mammograms,” IEE Conf. Publ., no. 409, pp. 442–447, 1995, doi: 10.1049/CP:19950597.

[4] P. Ardimento, L. Aversano, M. L. Bernardi, and M. Cimitile, “Deep Neural Networks Ensemble for Lung Nodule Detection on Chest CT Scans,” Proc. Int. Jt. Conf. Neural Networks, vol. 2021-July, Jul. 2021, doi: 10.1109/IJCNN52387.2021.9534176.

[5] M. Muzammil, I. Ali, I. U. Haq, A. A. Khaliq, and S. Abdullah, “Pulmonary nodule classification using feature and ensemble learning- based fusion techniques,” IEEE Access, vol. 9, pp. 113415–113427, 2021, doi: 10.1109/ACCESS.2021.3102707.

[6] I. Ali, M. Muzammil, I. U. Haq, A. A. Khaliq, and S. Abdullah, “Deep Feature Selection and Decision Level Fusion for Lungs Nodule Classification,” IEEE Access, vol. 9, pp. 18962–18973, 2021, doi: 10.1109/ACCESS.2021.3054735.

[7] D. Sethi, K. Arora, and S. Susan, “Transfer Learning by Deep Tuning of Pre-trained Networks for Pulmonary Nodule Detection,” 2020 IEEE 15th Int. Conf. Ind. Inf. Syst. ICIIS 2020 - Proc., pp. 168–173, Nov. 2020, doi: 10.1109/ICIIS51140.2020.9342686.

[8] R. Tekade and K. Rajeswari, “Lung Cancer Detection and Classification Using Deep Learning,” Proc. - 2018 4th Int. Conf. Comput. Commun. Control Autom. ICCUBEA 2018, Jul. 2018, doi: 10.1109/ICCUBEA.2018.8697352.

[9] M. S. Rahman, P. C. Shill, and Z. Homayra, “A New Method for Lung Nodule Detection Using Deep Neural Networks for CT Images,” 2nd Int. Conf. Electr. Comput. Commun. Eng. ECCE 2019, Apr. 2019, doi: 10.1109/ECACE.2019.8679439.

[10] S. K. Fazilov, O. R. Yusupov, and K. S. Abdiyeva, ‘‘Mammography image segmentation in breast cancer identification using the OTSU method,’’ Web. Sci., vol. 3, no. 8, pp. 196–205, Aug. 2022.

[11] K. V. Shiny and N. Sugitha, ‘‘Automatic brain tumor segmentation on pre-operative MRI using region growing algorithm,’’ in Information and Communication Technology for Competitive Strategies (ICTCS 2021). Cham, Switzerland: Springer, 2022, pp. 653–665.

[12] C. J. J. Sheela and G. Suganthi, ‘‘Morphological edge detection and brain tumor segmentation in magnetic resonance (MR) images based on region growing and performance evaluation of modified fuzzy C-means (FCM) algorithm,’’ Multimedia Tools Appl., vol. 79, nos. 25–26, pp. 17483–17496, Jul. 2020.

[13] H. Deng, J. P. Fitts, and C. A. Peters, ‘‘Quantifying fracture geometry with X-ray tomography: Technique of iterative local thresholding (TILT) for 3D image segmentation,’’ Comput. Geosci., vol. 20, no. 1, pp. 231–244, Feb. 2016.

[14] Y. Akbari, H. Hassen, S. Al-Maadeed, and S. M. Zughaier, ‘‘COVID-19 lesion segmentation using lung CT scan images: Comparative study based on active contour models,’’ Appl. Sci., vol. 11, no. 17, p. 8039, Aug. 2021.

[15] P. M. Bruntha, D. J. Rose, A. T. Shruthi, K. G. Juliet, and M. Kanimozhi, ‘‘Application of selective region growing algorithm in lung nodule segmentation,’’ in Proc. 4th Int. Conf. Devices, Circuits Syst. (ICDCS), Mar. 2018, pp. 319–322.

[16] P. B. Sangamithraa and S. Govindaraju, ‘‘Lung tumour detection and classification using EK-mean clustering,’’ in Proc. Int. Conf. Wireless Commun., Signal Process. Netw. (WiSPNET), Mar. 2016, pp. 2201–2206.

[17] M. Keshani, Z. Azimifar, F. Tajeripour, and R. Boostani, ‘‘Lung nodule segmentation and recognition using SVM classifier and active contour modeling: A complete intelligent system,’’ Comput. Biol. Med., vol. 43, no. 4, pp. 287–300, May 2013.

[18] W. J. Kostis, A. P. Reeves, D. F. Yankelevitz, and C. I. Henschke, ‘‘Threedimensional segmentation and growth-rate estimation of small pulmonary nodules in helical CT images,’’ IEEE Trans. Med. Imag., vol. 22, no. 10, pp. 1259–1274, Oct. 2003.

[19] X. Hongtao, D. Yang, N. Sun, Z. Chen, and Y. Zhang, ‘‘Automated pulmonary nodule detection in CT images using deep convolutional neural networks,’’ Pattern Recognit., vol. 85, pp. 109–119, Jan. 2019.

[20] J. Long, E. Shelhamer, and T. Darrell, ‘‘Fully convolutional networks for semantic segmentation,’’ in Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR), Jun. 2015, pp. 3431–3440.

[21] O. Ronneberger, P. Fischer, and T. Brox, ‘‘U-Net: Convolutional networks for biomedical image segmentation,’’ in Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent., 2015, pp. 234–241.

[22] L.-C. Chen, G. Papandreou, F. Schroff, and H. Adam, ‘‘Rethinking atrous convolution for semantic image segmentation,’’ 2017, arXiv:1706.05587.

[23] X. Luo, T. Song, G. Wang, J. Chen, Y. Chen, K. Li, D. N. Metaxas, and S. Zhang, ‘‘SCPM-Net: An anchor-free 3D lung nodule detection network using sphere representation and center points matching,’’ Med. Image Anal., vol. 75, Jan. 2022, Art. no. 102287.

[24] Z. Gu, J. Cheng, H. Fu, K. Zhou, H. Hao, Y. Zhao, T. Zhang, S. Gao, and J. Liu, ‘‘Ce-Net: Context encoder network for 2D medical image segmentation,’’ IEEE Trans. Med. Imag., vol. 38, no. 10, pp. 2281–2292, Oct. 2019.

[25] S. Bose, R. S. Chowdhury, R. Das, and U. Maulik, ‘‘Dense dilated deep multiscale supervised U-network for biomedical image segmentation,’’ Comput. Biol. Med., vol. 143, Apr. 2022, Art. no. 105274.

[26] Q. Dou, L. Yu, H. Chen, Y. Jin, X. Yang, J. Qin, and P.-A. Heng, ‘‘3D deeply supervised network for automated segmentation of volumetric medical images,’’ Med. Image Anal., vol. 41, pp. 40–54, Oct. 2017.

[27] F. Milletari, N. Navab, and S.-A. Ahmadi, ‘‘V-Net: Fully convolutional neural networks for volumetric medical image segmentation,’’ in Proc. 4th Int. Conf. 3D Vis. (3DV), Oct. 2016, pp. 565–571.

[28] K. He, X. Zhang, S. Ren, and J. Sun, ‘‘Deep residual learning for image recognition,’’ in Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR), Jun. 2016, pp. 770–778.

[29] Y. Li and Y. Fan, ‘‘DeepSEED: 3D Squeeze-and-excitation encoder– decoder convolutional neural networks for

pulmonary nodule detection,'' in Proc. IEEE 17th Int. Symp. Biomed. Imag. (ISBI), Apr. 2020, pp. 1866–1869.

[30] J. Deng, J. Guo, N. Xue, and S. Zafeiriou, ''ArcFace: Additive angular margin loss for deep face recognition,'' in Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR), Jun. 2019, pp. 4690–4699.

[31] S. G. Armato III et al., ''The lung image database consortium (LIDC) and image database resource initiative (IDRI): A completed reference database of lung nodules on CT scans,'' Med. Phys., vol. 38, no. 2, pp. 915–931, 2011.

[32] A. A. A. Setio et al., ''Validation, comparison, and combination of algorithms for automatic detection of pulmonary nodules in computed tomography images: The LUNA16 challenge,'' Med. Image Anal., vol. 42, pp. 1–13, Dec. 2017.

[33] Z. Zhou, M. M. R. Siddiquee, N. Tajbakhsh, and J. Liang, ''UNet++: A nested U-Net architecture for medical image segmentation'', Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support. Cham, Switzerland: Springer, 2018, pp. 3–11.

[34] Jiachen Hou et. al., "Lung Nodule Segmentation Algorithm with SMR-Unet", IEEE Access, April 2023.

[35] Weilun Wang et. al., "3D Multiscale DenseNet for Malignancy Grade Classification of Pulmonary Nodules", International Conference on Awareness Science and Technology (ICAST), 2020.

[36] https://radiologyassistant.nl/chest/lung-cancer/tnm-classification-8th-edition

[37] Sobia Shafiq et. al., "An Effective Early Stage Detection of Lung Cancer using Fuzzy Local Information cMean and GoogleLeNet", 16th International Conference on Open Source Systems and Technologies (ICOSST), 2022.

[38] Mohd Firdaus Abdullah et. al., "Classification of Lung Cancer Stages from CT Scan Images Using Image Processing and k-Nearest Neighbours", 11th IEEE Control and System Graduate Research Colloquium (ICSGRC), 2020.

[39] Deepika Gupta et. al., "Detection and Staging of Lung Cancer from CT scan Images by Deep Learning", International Conference on Disruptive Technologies (ICDT), 2023.

[40] F. Liao, M. Liang, Z. Li, X. Hu, and S. Song, "Evaluate the Malignancy of Pulmonary Nodules Using the 3-D Deep Leaky Noisy-OR Network", IEEE Trans. Neural Networks Learn. Syst., vol. 30, no. 11, pp. 3484–3495, 2019, doi: 10.1109/TNNLS.2019.2892409.

[41] C. Shorten and T. M. Khoshgoftaar, ''A survey on image data augmentation for deep learning,'' J. Big Data, vol. 6, no. 1, p. 60, 2019, doi: 10.1186/s40537-019-0197-0.